

Palindromic richness and Coxeter groups

Edita Pelantová

Czech Technical University in Prague
Czech Republic

Štěpán Starosta

Czech Technical University in Prague
Czech Republic

Abstract

For a given finite group G consisting of morphisms and antimorphisms of a free monoid \mathcal{A}^* , we study infinite words with language closed under the group G . We focus on words rich in generalized palindromic factors, i.e., in factors w satisfying $\Theta(w) = w$ for some non-identical element $\Theta \in G$. We give several equivalent descriptions of rich words and show the role Coxeter groups play in the generalized notion of palindromic richness.

1 Introduction

In [15], Droubay et al. showed that the number of different palindromes occurring in a finite word w cannot exceed the bound $|w| + 1$. If this bound is met, the word w is called rich or rich in palindromes or full [15, 7]. An infinite word \mathbf{u} is said to be rich if any its factor w is rich. The list of the most prominent rich words contains Arnoux-Rauzy words and words coding interval exchange with symmetric permutation of intervals (note that Sturmian words belong to both mentioned groups of words).

During the past two decades, the notion of *palindromic richness* showed to be fruitful. Application of *palindromes* in physics of quasicrystals (see for instance [17, 13]) and in genetics (see for instance [20]) served as stimulating factor for research in this area as well. Recently Restivo and Rosone [25] showed a narrow connection of rich periodic words with extremal cases of Burrows-Wheeler transform which is used in compression algorithms.

Generalizations of rich words appeared soon. Instead of classical palindromes defined as words invariant under the reversal mapping, one considers Θ -*palindromes*, i.e., words invariant under an involutive antimorphism Θ . For first appearance of the notion see [19], where it appeared in the context of DNA, or [14], where the name pseudopalindrome is also used. Words saturated by Θ -palindromes up to the highest possible level are called Θ -*rich*. Another kind of generalization of rich words relaxes the requirement on

the number of palindromes occurring in any factor w . We say that infinite word \mathbf{u} is *almost rich*, if there exists a constant D such that any factor w of \mathbf{u} contains at least $|w| + 1 - D$ different palindromes. The minimal constant D with this property is referred to as *palindromic defect* and was introduced in [7]. Both mentioned generalizations can be combined into the notion of *almost Θ -rich* words. It follows directly from the definition that almost Θ -rich words contain infinitely many Θ -palindromes.

As shown in [23], besides peculiar periodic words, no infinite uniformly recurrent words can be simultaneously almost Θ_1 -rich and almost Θ_2 -rich for two different involutive antimorphisms Θ_1 and Θ_2 . The famous Thue-Morse word contains infinitely many classical palindromes and Θ -palindromes, where Θ is antimorphism generated by interchange of symbols 0 and 1. Nevertheless, the Thue-Morse word has no chance to be almost rich or almost Θ -rich. Therefore, the authors have suggested in [23] a further generalization under the name of *G -richness*. The new definition of richness respects all antimorphisms of finite order under which the language of an infinite word is invariant, the letter G stands for the group generated by these antimorphisms. The definition is based on the notion of *graph of symmetries*, which is assigned to an infinite word. Adopting the new definition, the second author showed in [26] that all generalized Thue-Morse words $t_{b,m}$ are D_m -rich, where D_m is a group isomorphic to the dihedral group having $2m$ elements.

It turned out that words rich in the classical sense can be characterized by using many other notions of combinatorics on words: return word, bilateral order, longest palindromic suffix, factor and palindromic complexity, and Rauzy graphs. These characterizations can be found in [16, 10, 3]. Analogous results for Θ -rich words can be found in [27]. The aim of this article is to find G -analogies of these characterizations. They are stated as Theorems 13, 20 and 31 and Proposition 35. An important consequence of these characterizations is the fact that existence of a G -rich word forces the group G to be a finite Coxeter group, i.e., a group generated by involutive elements only, for definition see for instance [18]. The question whether for any finite Coxeter group G there exists a G -rich word remains open. At the end of the article we present two examples of G -rich word, in which G is a Coxeter group not isomorphic to a dihedral group.

2 Preliminaries

An *alphabet* \mathcal{A} is a finite set. Elements of \mathcal{A} are usually called *letters*. A *finite word* w over \mathcal{A} is a finite string $w = w_1 w_2 \dots w_n$ of letters $w_i \in \mathcal{A}$. Its length, denoted by $|w|$, equals n . The set of all finite words over \mathcal{A} equipped with the operation of concatenation is the free monoid \mathcal{A}^* . Its neutral element is the *empty word* ε . A word $v \in \mathcal{A}^*$ is a *factor* of a word $w \in \mathcal{A}^*$ if there exist words $s, t \in \mathcal{A}^*$ such that $w = svt$. If $s = \varepsilon$, then v is a *prefix* of w , if $t = \varepsilon$, then v is a *suffix* of w .

2.1 Antimorphisms and their fixed points

A mapping φ on \mathcal{A}^* is called

- a *morphism* if $\varphi(vw) = \varphi(v)\varphi(w)$ for any $v, w \in \mathcal{A}^*$;
- an *antimorphism* if $\varphi(vw) = \varphi(w)\varphi(v)$ for any $v, w \in \mathcal{A}^*$.

We denote the set of all morphisms and antimorphisms on \mathcal{A}^* by $AM(\mathcal{A}^*)$. Together with composition, it forms a monoid with the identity mapping Id as the unit element. The set of all morphisms, denoted by $M(\mathcal{A}^*)$, is a submonoid of $AM(\mathcal{A}^*)$. The *reversal* mapping R defined by

$$R(w_1w_2 \dots w_n) = w_nw_{n-1} \dots w_2w_1$$

is an involutive antimorphism, i.e., $R^2 = \text{Id}$. It is obvious that any antimorphism is a composition of R and a morphism. Thus

$$AM(\mathcal{A}^*) = M(\mathcal{A}^*) \cup R(M(\mathcal{A}^*)).$$

A fixed point of a given antimorphism Θ is called Θ -*palindrome*, i.e., a word w is a Θ -palindrome if $w = \Theta(w)$. If Θ is the reversal mapping R , we say *palindrome* or *classical palindrome* instead of R -palindrome. One can see that if Θ has a fixed point containing all the letters of \mathcal{A} , then Θ is an involution, and thus a composition of R and an involutive permutation of letters.

2.2 Factor and palindromic complexities

An *infinite word* \mathbf{u} over an alphabet \mathcal{A} is a sequence $(u_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$. We always implicitly suppose that \mathcal{A} is the smallest possible alphabet for \mathbf{u} , i.e., any letter of \mathcal{A} occurs at least once in \mathbf{u} . Action of any morphism $\varphi \in M(\mathcal{A}^*)$ can be naturally extended to infinite words by the prescription

$$\varphi(\mathbf{u}) = \varphi(u_0)\varphi(u_1)\varphi(u_2) \dots \quad \text{for all } \mathbf{u} = (u_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}.$$

A finite word w is a *factor* of \mathbf{u} if there exists an index $i \in \mathbb{N}$, called *occurrence* of w , such that $w = u_iu_{i+1} \dots u_{i+|w|-1}$. The set of all factors of \mathbf{u} of length n is denoted $\mathcal{L}_n(\mathbf{u})$. The *language* of an infinite word \mathbf{u} is the set of all its factors $\mathcal{L}(\mathbf{u}) = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n(\mathbf{u})$. An infinite word \mathbf{u} is *recurrent* if any its factor has at least two occurrences in \mathbf{u} . A factor $v \in \mathcal{L}(\mathbf{u})$ is a *complete return word* of a factor w if w occurs in v exactly twice, as a suffix and a prefix of v . A complete return word v of w can be written as $v = qw$ for some factor $q \neq \varepsilon$, which is usually called a *return word* of w . If any factor w of \mathbf{u} has only finitely many return words, then \mathbf{u} is said to be *uniformly recurrent*.

The *factor complexity* of \mathbf{u} is the mapping $\mathcal{C} : \mathbb{N} \mapsto \mathbb{N}$ defined by the prescription

$$\mathcal{C}(n) := \#\mathcal{L}_n(\mathbf{u}).$$

To evaluate the factor complexity of \mathbf{u} , one may watch possible prolongations of factors. A letter $a \in \mathcal{A}$ is a *left extension* of a factor w in \mathbf{u} if aw belongs to $\mathcal{L}(\mathbf{u})$. The set of all left extensions of w is denoted $\text{Lext}(w)$. A factor $w \in \mathcal{L}(\mathbf{u})$ is called *left special*, if $\#\text{Lext}(w) \geq 2$. Analogously, we define *right extension*, the set $\text{Rext}(w)$, and *right*

special. If w is right and left special, we call it *bispecial*. The first difference of the factor complexity of a recurrent word \mathbf{u} satisfies

$$\Delta\mathcal{C}(n) = \mathcal{C}(n+1) - \mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} (\#\text{Lext}(w) - 1) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} (\#\text{Rext}(w) - 1).$$

The second difference of factor complexity can be expressed using the *bilateral order* of a factor. Let w be a factor of \mathbf{u} . Its bilateral order is the quantity $b(w) := \#\{awb \mid awb \in \mathcal{L}(\mathbf{u}), a, b \in \mathcal{A}\} - \#\text{Lext}(w) - \#\text{Rext}(w) + 1$. In [12], the formula

$$\Delta^2\mathcal{C}(n) = \Delta\mathcal{C}(n+1) - \Delta\mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} b(w)$$

is deduced.

The Θ -palindromic complexity of \mathbf{u} is the mapping $\mathcal{P}_\Theta(n) : \mathbb{N} \mapsto \mathbb{N}$ defined by

$$\mathcal{P}_\Theta(n) := \#\{w \in \mathcal{L}_n(\mathbf{u}) \mid w = \Theta(w)\}.$$

Similarly to factor complexity, evaluation of palindromic complexity is based on counting possible extensions of palindromes. For a Θ -palindrome w , we denote by $\text{Pext}_\Theta(w)$ the set of all letters $a \in \mathcal{A}$ such that $aw\Theta(a) \in \mathcal{L}(\mathbf{u})$. It is easy to see that

$$\mathcal{P}_\Theta(n+2) - \mathcal{P}_\Theta(n) = \sum_{\substack{w \in \mathcal{L}_n(\mathbf{u}) \\ w = \Theta(w)}} (\#\text{Pext}_\Theta(w) - 1). \quad (1)$$

2.3 Words with language closed under a group G and G -richness

In the sequel, the symbol G stands exclusively for a finite subgroup of $AM(\mathcal{A}^*)$ containing at least one antimorphism. The set of all involutive antimorphisms belonging to G is denoted by $G^{(2)}$. We say that finite words $w, v \in \mathcal{A}^*$ are G -equivalent if there exists $\mu \in G$ such that $w = \mu(v)$. The class of equivalence containing a word w is denoted

$$[w] := \{\mu(w) \mid \mu \in G\}.$$

Since the group G is finite, any $\mu \in G$ preserves length of words and thus equivalent words have the same length.

We say that language $\mathcal{L}(\mathbf{u})$ of an infinite word $\mathbf{u} \in \mathcal{A}^\mathbb{N}$ is *closed under G* if for any factor $w \in \mathcal{L}(\mathbf{u})$ and any $\mu \in G$, the image $\mu(w)$ belongs to $\mathcal{L}(\mathbf{u})$ as well. Since G contains at least one antimorphism, closedness of $\mathcal{L}(\mathbf{u})$ under G implies that \mathbf{u} is recurrent.

The following graph assigned to such a word \mathbf{u} was defined in [23]. It generalizes the notion of super reduced Rauzy graph introduced in [10].

Definition 1. Let \mathbf{u} be an infinite word with language closed under G and $n \in \mathbb{N}$.

- 1) The directed graph of symmetries of the word \mathbf{u} is $\vec{\Gamma}_n(\mathbf{u}) = (V, \vec{E})$ with the set of vertices

$$V = \{[w] \mid w \in \mathcal{L}_n(\mathbf{u}), w \text{ is left or right special}\}$$

and an edge $e \in \vec{E} \subset \mathcal{L}(\mathbf{u})$ starts in the vertex $[w]$ and ends in the vertex $[v]$, if

- the prefix of e of length n belongs to $[w]$,
- the suffix of e of length n belongs to $[v]$,
- e has exactly two occurrences of special factors of length n .

- 2) The undirected graph of symmetries of the word \mathbf{u} is $\Gamma_n(\mathbf{u}) = (V, E)$ with the same set of vertices as $\vec{\Gamma}_n(\mathbf{u})$ and for any $e \in \mathcal{L}(\mathbf{u})$ we have

$$[e] \in E \iff e \in \vec{E}.$$

Any vertex $[w]$ of the graph $\vec{\Gamma}_n(\mathbf{u})$ represents a class of equivalent factors of $\mathcal{L}_n(\mathbf{u})$. It has at most $\#G$ elements; the actual cardinality of $[w]$ may depend on n as well.

Definition 2. Let \mathbf{u} be an infinite word with language closed under G . We say that a number $n \in \mathbb{N}$ is G -distinguishing on \mathbf{u} if for any $w \in \mathcal{L}_n(\mathbf{u})$ it holds:

$$\Theta_1 \neq \Theta_2 \implies \Theta_1(w) \neq \Theta_2(w) \quad \text{for any two antimorphisms } \Theta_1, \Theta_2 \in G. \quad (2)$$

If n is G -distinguishing on \mathbf{u} , then the knowledge of a pair w and $\Theta(w)$ for a single word $w \in \mathcal{L}_n(\mathbf{u})$ enables us to determine Θ . Let us stress that the requirement (2) gives also for any two distinct morphisms $\phi_1, \phi_2 \in G$ that $\phi_1(w) \neq \phi_2(w)$. Thus, n is the length of words on which action of any element of the group G is fully demonstrated.

One can readily see that if n is G -distinguishing on \mathbf{u} , then any $m > n$ is also G -distinguishing on \mathbf{u} .

As proved in [23], connectivity of graphs of symmetries enables to bind the factor and palindromic complexities.

Theorem 3. Let \mathbf{u} be an infinite word with language is closed under G and $N \in \mathbb{N}$ be G -distinguishing on \mathbf{u} . Then

$$\Delta\mathcal{C}(n) + \#G \geq \sum_{\Theta \in G^{(2)}} \left(\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n+1) \right) \quad \text{for any } n \geq N. \quad (3)$$

The term $\Delta\mathcal{C}(n) + \#G$ represents an upper bound on the number of palindromes occurring in \mathbf{u} . Words for which this upper bound is reached are in some sense opulent in palindromes. Proof of Theorem 3 suggested us to adopt the following definition of Property G -graph(N).

Definition 4. We say that an infinite word \mathbf{u} has Property G -graph(N) if for each $n \in \mathbb{N}$, $n \geq N$ it holds

- $\mathcal{L}(\mathbf{u})$ is closed under G ;
- if $[e]$ is a loop in $\Gamma_n(\mathbf{u})$, then e is a Θ -palindrome for some $\Theta \in G$;
- the graph obtained from $\Gamma_n(\mathbf{u})$ by removing loops is a tree.

Definition 5. We say that \mathbf{u} is G -rich if \mathbf{u} has Property G -graph(1) and \mathbf{u} is almost G -rich if there exists $N \in \mathbb{N}$ such that \mathbf{u} has Property G -graph(N).

In the next section, we explain legitimacy of the name G -richness, i.e., we show that the classical richness is contained in our new definition of richness as well.

2.4 Palindromic richness in the classical sense

Let us recall the origin of palindromic richness. In this section, we use the word palindrome for R -palindrome where R denotes the reversal mapping, and we denote by $\text{Pal}(w)$ the set of all palindromic factors of a word w including the empty word ε . In [15], Droubay, Justin and Pirillo provided the simple upper bound

$$\#\text{Pal}(w) \leq |w| + 1. \quad (4)$$

This bound serves for the definition of richness in the classical sense.

Definition 6. A finite word w is called rich if the equality in (4) holds. An infinite word is rich if all its factors are rich.

Another type of bound on number of palindromes contained in an infinite word was proved in [1]. If an infinite word has its language closed under the reversal mapping, then the following inequality holds

$$\Delta\mathcal{C}(n) + 2 \geq \mathcal{P}(n) + \mathcal{P}(n+1) \quad \text{for all } n \in \mathbb{N}. \quad (5)$$

The authors of [16] showed that an infinite word \mathbf{u} with language closed under reversal is rich if and only if the equality in (5) is attained for all $n \in \mathbb{N}$. Their proof uses the notion of super reduced Rauzy graph, which, in our terminology, is the graph of symmetries $\Gamma_n(\mathbf{u})$ for the group $G = \{\text{Id}, R\}$. A word \mathbf{u} is shown to be rich if and only if \mathbf{u} satisfies - again in our terminology - Property G -graph(1).

In Section 6, we illustrate on the Thue-Morse word why Definition 5 of G -richness is based on Property G -graph(1) instead of a modification of the inequality (4).

Let us point out a drawback of our new definition. The Property G -graph(1), unlike the classical Definition 6 of richness, requires $\mathcal{L}(\mathbf{u})$ to be closed under reversal. Nevertheless, an infinite word can be rich in the classical sense without having its language closed under reversal. On the other hand, as proved in [16] (Proposition 2.11), any recurrent rich word has its language closed under reversal. Therefore, on the set of recurrent words both definitions coincide. The following theorem summarizes properties characterizing rich recurrent words, their proofs can be found in [16, 15, 10, 3].

Theorem 7. *For an infinite word \mathbf{u} with language closed under reversal the following statements are equivalent:*

1. \mathbf{u} is rich,
2. any complete return word of any palindromic factor of \mathbf{u} is a palindrome,
3. for any factor w of \mathbf{u} , every factor of \mathbf{u} that contains w only as its prefix and $R(w)$ only as its suffix is a palindrome,
4. the longest palindromic suffix of any factor $w \in \mathcal{L}(\mathbf{u})$ is unioccurrent in w ,
5. for each n the following equality holds

$$\mathcal{C}(n+1) - \mathcal{C}(n) + 2 = \mathcal{P}(n) + \mathcal{P}(n+1),$$

6. each graph of symmetries $\Gamma_n(\mathbf{u})$ satisfies: all its loops are palindromes and the graph obtained from $\Gamma_n(\mathbf{u})$ by removing loops is a tree,
7. any bispecial factor w of \mathbf{u} satisfies:

- if w is non-palindromic, then

$$\mathbf{b}(w) = 0,$$

- if w is a palindrome, then

$$\mathbf{b}(w) = \#\text{Pext}(w) - 1.$$

Another characterization of rich words can be found in [11]. Almost rich words were introduced in [16]. The definition uses the notion of defect of an infinite word, which is based on the inequality (4). Therefore, we once more postpone the discussion about a relation between the notion of almost G -almost richness and almost richness to Section 6. Words almost rich in the classical sense can be characterized by properties analogous to those listed in Theorem 7. For more details see [4].

As already mentioned, in the articles [19, 14], the reversal mapping R is replaced by an arbitrary involutive antimorphism Θ and Θ -palindrome is defined as a word v satisfying $\Theta(v) = v$. Let us denote by $\text{Pal}_\Theta(w)$ the set of Θ -palindromic factors occurring in w . As shown in [27],

$$\#\text{Pal}_\Theta(w) \leq |w| + 1 - \gamma_\Theta(w), \quad (6)$$

where $\gamma_\Theta(w) := \#\{\{a, \Theta(a)\} \mid a \in \mathcal{A}, a \text{ occurs in } w \text{ and } a \neq \Theta(a)\}$. Analogously to the classical richness, Θ -richness is introduced. A finite word w is called Θ -rich if the equality in (6) holds. An infinite word is Θ -rich if all its factors are Θ -rich. Generalizations of some characterizations in Theorem 7 for Θ -rich words are presented in [27] as well.

3 Tools for characterization of G -rich words

Classical rich and almost rich words can be described using the notions of return words and longest palindromic suffix. In order to find a suitable description of G -richness, we at first introduce G -analogies of these notions and in the sequel we demonstrate their efficiency. Let us recall that in all definitions and statements in the sequel, the symbol G stands for a finite subgroup of $AM(\mathcal{A}^*)$ which contains at least one antimorphism.

Definition 8. A word $w \in \mathcal{A}^*$ is said to be G -palindrome if there exists an antimorphism $\Theta \in G$ such that $w = \Theta(w)$.

Definition 9. Let $w, v \in \mathcal{A}^*$. G -occurrence of a word w in a word v is an index i such that there exists $w' \in [w]$ having occurrence i in v .

We say that w is G -unioccurrent in v if w occurs in v and there is no other G -occurrence of w in v .

Definition 10. Let $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ be an infinite word and $w \in \mathcal{L}(\mathbf{u})$. A factor $v \in \mathcal{L}(\mathbf{u})$ of length $|v| > |w|$ is called complete G -return word of $[w]$ in \mathbf{u} if

- a prefix and a suffix of v belong to $[w]$ and
- v contains no other occurrence of a word from $[w]$.

We say that $v' \in \mathcal{L}(\mathbf{u})$ is a G -return word of $[w]$ in \mathbf{u} if for some $w' \in [w]$ the word $v'w'$ is a complete G -return word of $[w]$.

Definition 11. A suffix w of a word $v \in \mathcal{A}^*$ is called G -longest palindromic suffix of v if

- w is a G -palindrome and
- $|w| \geq |w'|$ for any G -palindromic suffix w' of v .

The G -longest palindromic suffix of v is denoted by $G\text{-lps}(v)$.

Remark 12. If no non-empty suffix w of a non-empty word $v \in \mathcal{A}^*$ is a G -palindrome, then $G\text{-lps}(v) = \varepsilon$. It may occur only in case when G does not contain the reversal mapping R . Let us recall that any index $i \in \{1, 2, \dots, n\}$ is defined to be an occurrence of ε in $v = v_1v_2 \dots v_{n-1}$.

4 G -richness and G -return words

In this section, we demonstrate that the notions of G -return word and G -palindrome can grasp the essence of G -richness as shown in the following theorem.

Theorem 13. Let \mathbf{u} be an infinite word with language closed under G . Then \mathbf{u} is G -rich if and only if for all $w \in \mathcal{L}(\mathbf{u})$ every complete G -return word of $[w]$ is a G -palindrome.

The statement for $G = \{\text{Id}, R\}$ is shown in [16] and for $G = \{\text{Id}, \Theta\}$ in [27]. A statement analogous to Theorem 13 holds for almost G -richness as well and we prove it in this subsection as well. Let us recall that according to Definition 5, almost G -richness means property $G\text{-graph}(N)$ for some N . First we introduce a new property based on G -return words which depends on a parameter N as well.

Definition 14. Let $N \in \mathbb{N}$. We say that $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ satisfies Property $G\text{-crw}(N)$ if for all $w \in \mathcal{L}(\mathbf{u})$, $|w| \geq N$, every complete G -return word of $[w]$ is a G -palindrome.

The proof of Theorem 13 is split into proofs of Lemmas 15 and 16 and Lemma 19.

Lemma 15. Let $N \in \mathbb{N}$ and $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ satisfy Property $G\text{-graph}(N)$. Let w be a factor of \mathbf{u} such that $|w| \geq N$ and v be a complete G -return word of $[w]$ in \mathbf{u} starting in w . Then there exist a letter $a \in \mathcal{A}$ and an antimorphism $\Theta \in G$ such that wa is a prefix of v and $\Theta(a)\Theta(w)$ is a suffix of v .

Proof. Take $n \geq N$, $w \in \mathcal{L}_n(\mathbf{u})$, and v a complete G -return word of $[w]$ starting in w . Denote $w' \in [w]$ the suffix of v of length n .

At first we suppose that w is a special factor of \mathbf{u} . We consider the following two cases. We exploit the graphs of symmetries $\vec{\Gamma}_n(\mathbf{u}) = (V, \vec{E})$ and $\Gamma_n(\mathbf{u}) = (V, E)$.

1. If besides the prefix w and the suffix w' the complete G -return word v contains no other occurrences of a special factor of length n , then $[v]$ is an edge in E which starts and ends in the same vertex $[w]$. The edge $[v]$ is thus a loop and according to the definition of Property $G\text{-graph}(N)$, the factor v is a G -palindrome. Obviously, v has the property stated in the claim.
2. Let $v = v_0v_1 \dots v_m$ contain a special factor $z \notin [w]$ of length n at the position i , i.e., $z = v_i \dots v_{i+n-1}$. We may suppose without loss of generality that i is the least index with this property. Then $[z]$ is a vertex of $\vec{\Gamma}_n(\mathbf{u})$ and a prefix of v is an edge in $\vec{\Gamma}_n(\mathbf{u})$ starting in $[w]$ and ending in $[z]$. Since the graph obtained from $\Gamma_n(\mathbf{u})$ by removing loops is a tree, the complete G -return word v has a suffix $f \in \vec{E}$ such that a prefix of f belongs to $[z]$ and its suffix belongs to $[w]$. Moreover, there exists an antimorphism Θ such that $f = \Theta(e)$. As $|e| = |f| > |w|$, the factor v has the property stated in the claim.

Let us now suppose that w is not a special factor and thus w has a unique right extension, say a . If $w' = \Theta(w)$ for some antimorphism $\Theta \in G$, then $\Theta(w)$ has a unique left extension $\Theta(a)$ and therefore $\Theta(a)\Theta(w)$ is a suffix of the complete G -return word v as stated in the claim.

To finish the proof, it is enough to consider the situation when $w' = \mu(w)$ for some morphism $\mu \in G$ and w is not a special factor. We discuss two separate cases.

1. There exists no special factor of length n .

In this case \mathbf{u} is periodic. Denote v' the word such that $v = v'\mu(w)$. Because no

special factor of length at least n exists, the factor $v = v'\mu(w)$ is the unique right prolongation of w of length $|v|$. As $\mathcal{L}(\mathbf{u})$ is closed under G , $\mu^\ell(v) = \mu^\ell(v')\mu^{\ell+1}(w)$ is the unique right prolongation of $\mu^\ell(w)$ of given length. In particular, for $\ell = 1$ it implies $v'\mu(v')\mu^2(w) \in \mathcal{L}(\mathbf{u})$. Repeating this argument for $\ell = 2, 3, \dots$, we deduce that $v'\mu(v')\mu^2(v') \dots \mu^\ell(v')\mu^{\ell+1}(w) \in \mathcal{L}(\mathbf{u})$. Therefore, the factor $v'\mu(v')\mu^2(v') \dots \mu^{k-1}(v')$, where k is the order of the morphism μ , is a period of \mathbf{u} which does not contain any antimorphic image of w - a contradiction.

2. There exists a special factor of length n .

Consequently, there exists a unique q such that wq is right special and no proper prefix of wq is right special. The factor wq has only one occurrence of a factor $\nu(w)$ for some morphism $\nu \in G$ - in the opposite case, we can find a shorter prolongation of w which is right special. Since v has suffix $\mu(w)$, we deduce $|wq| < |v|$. As μ is a morphism, $\mu(w)\mu(q)$ is the only right prolongation of $\mu(w)$ and thus $v\mu(q)$ is a complete G -return word of $[wq]$. For the special factor wq , we may now use the first part of the proof and thus find an antimorphism Θ such that $\mu(w)\mu(q) = \Theta(wq) = \Theta(q)\Theta(w)$. Applying morphism μ^{-1} , we get $wq = \mu^{-1}\Theta(q)\mu^{-1}\Theta(w)$. Together with the inequality $|wq| < |v|$, it implies a contradiction with the fact that v is a G return word of $[w]$.

□

Lemma 16. *Let $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ and $N \in \mathbb{N}$. If \mathbf{u} has Property $G\text{-graph}(N)$, then it has Property $G\text{-crw}(N)$.*

Proof. Let v be a complete G -return word of $[w]$ starting in w for a factor w such that $|w| \geq N$. Denote by a a letter such that wa is a prefix of v .

If $wa = v$, then Lemma 15 implies that v is a G -palindrome.

If $wa \neq v$, then according to Lemma 15, v is a complete G -return word of $[wa]$ as well. We apply the procedure again on wa . We find a letter b such that wab is a prefix of v . If $wab = v$, the v is a G -palindrome, otherwise v is a complete G -return word of $[wab]$.

We continue in this way until the procedure stops and we conclude that v is a G -palindrome. □

Remark 17. For an infinite word \mathbf{u} satisfying Property $G\text{-graph}(N)$, the occurrences of morphic and antimorphic images of w , $|w| \geq N$, alternate. This consequence of the previous claim is an analogy to the claims stated in [16] for rich words and in [27] for Θ -rich words.

Theorem 18. *Let there exist an almost G -rich word. Then G is a Coxeter group.*

Proof. At first suppose that the G -rich word \mathbf{u} is periodic. According to [23], for a periodic word \mathbf{u} the closedness of $\mathcal{L}(\mathbf{u})$ under an antimorphism Θ forces Θ to be involutive. Thus, all antimorphisms in G are involutive and therefore G is a Coxeter group.

Now suppose that the G -rich word \mathbf{u} is aperiodic. Let w be a prefix of a G -rich word \mathbf{u} such that any letter of the alphabet occurs in w . According to Lemma 15, any occurrence

of a word from $[w]$ is a Θ -image of the left closest occurrence of a factor from $[w]$, where Θ is an antimorphism. Since all letters occur in w , the antimorphism Θ is involutive. Therefore, any factor $\nu(w)$ occurring in \mathbf{u} can be written as $\nu(w) = \Theta_1\Theta_2\ldots\Theta_s(w)$, where $\Theta_1\Theta_2\ldots\Theta_s$ is a sequence of involutive antimorphisms. As w contains all letters, the equality $\nu(w) = \Theta_1\Theta_2\ldots\Theta_s(w)$ implies $\nu = \Theta_1\Theta_2\ldots\Theta_s$. □

Lemma 19. *Let \mathbf{u} be an infinite word with language closed under G and $N \in \mathbb{N}$. If \mathbf{u} satisfies Property $G\text{-crw}(N)$, then it satisfies Property $G\text{-graph}(N)$.*

Proof. Let $n \geq N$ and $w \in \mathcal{L}_n(\mathbf{u})$. We assume that every complete G -return word of $[w]$ is a G -palindrome. We have to show two properties of $\Gamma_n(\mathbf{u})$.

1. Any loop in $\Gamma_n(\mathbf{u})$ is a G -palindrome.
 Since any loop e in $\Gamma_n(\mathbf{u})$ at a vertex $[w]$ is a complete G -return word of $[w]$, the loop e is a G -palindrome by our assumption.
2. The graph obtained from $\Gamma_n(\mathbf{u})$ by removing loops is a tree.
 Or equivalently, we show that in $\Gamma_n(\mathbf{u})$ there exists unique path between any two different vertices $[w']$ and $[w'']$. Let p be a factor of \mathbf{u} such that a prefix of p belongs to $[w']$, its suffix belongs to $[w'']$ and p has no other occurrences of factor from $[w']$ or $[w'']$. Let without loss of generality w' be a prefix of p . Let us find a complete G -return word of $[w']$ with prefix p , denote it v . Since v is a G -palindrome, the factor $\Theta(p)$ is a suffix of v for some antimorphism $\Theta \in G$. As v is a complete G -return word of $[w']$, v has exactly two G -occurrences of w' . The factor v contains at least two G -occurrences of w'' . Therefore, the next factor with the same properties as p , i.e., representing a path in the undirected graph $\Gamma_n(\mathbf{u})$ between $[w']$ and $[w'']$, which occurs in \mathbf{u} after p , is $\Theta(p)$. Consequently, any factor with the same properties as p belongs to the same equivalence class $[p]$. □

5 G -richness and G -longest palindromic suffix

As we have already stated, the classical richness is connected to the number of occurrences of the longest palindromic suffix in any factor. In the case of classical palindromes, the longest palindromic suffix of any word is always non-empty, but it is not always satisfied for the G -longest palindromic suffix. Therefore, the characterization of G -richness by the G -longest palindromic suffix needs a modification. The classical case of the following statement is established in [16] and for Θ -richness in [27].

Theorem 20. *Let $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ be an infinite word with language closed under G . Then \mathbf{u} is G -rich if and only if for any factor $v \in \mathcal{L}(\mathbf{u})$, its G -longest palindromic suffix is G -unioccurrent in v or the last letter of v is G -unioccurrent in v .*

Again we prove a more general statement, since an analogous theorem holds for almost G -richness as well. At first we introduce a new property depending on a parameter N .

Definition 21. Let $N \in \mathbb{N}$. We say that $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ satisfies Property $G\text{-lps}(N)$ if for all $w \in \mathcal{L}(\mathbf{u})$, $|w| \geq N$, either the word $G\text{-lps}(w)$ is G -unioccurrent in w , or the suffix of w of length 1 has exactly one G -occurrence in w .

Lemma 22. Let $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$.

1. If \mathbf{u} has Property $G\text{-crw}(1)$, then \mathbf{u} has Property $G\text{-lps}(1)$.
2. Let $N \in \mathbb{N}$. If \mathbf{u} is uniformly recurrent and has Property $G\text{-crw}(N)$, then there exists $M \in \mathbb{N}$ such that \mathbf{u} has Property $G\text{-lps}(M)$.

Proof. Let us realize a trivial fact: if $w \in \mathcal{L}(\mathbf{u})$ has a suffix v which is not G -unioccurrent in w , then there is a suffix of w which is a complete G -return word of $[v]$.

To prove the first assertion, consider a factor $w \in \mathcal{L}(\mathbf{u})$. If the last letter of w , denoted a , is G -unioccurrent, we have nothing to do. If a is not G -unioccurrent in w , then according to the mentioned fact and by Property $G\text{-crw}(1)$, a complete G -return word of $[a]$ is a G -palindrome of length greater than 1. Therefore, $G\text{-lps}(w) \neq \varepsilon$. We have to show that $G\text{-lps}(w)$ is G -unioccurrent in w . If not, then a suffix of w is a complete G -return word of $[G\text{-lps}(w)]$ which, according to Property $G\text{-crw}(1)$, is a G -palindrome longer than the G -longest palindromic suffix of w - a contradiction.

Now we prove the second assertion. Since \mathbf{u} is uniformly recurrent, there exists an integer M such that every factor $w \in \mathcal{L}(\mathbf{u})$, $|w| \geq M$, contains at least two occurrences of every factor $v \in \mathcal{L}_N(\mathbf{u})$. It particular, it implies that w contains at least two G -occurrences of its suffix z of length N . By the fact mentioned at the beginning of the proof, Property $G\text{-crw}(N)$ implies that $G\text{-lps}(w)$ is longer than N and is G -unioccurrent in w . □

Remark 23. Let us note that the second part of the previous lemma can be proved considering a weaker assumption than uniform recurrence of \mathbf{u} . It is enough to assume that every factor $w \in \mathcal{L}(\mathbf{u})$ has only finitely many complete G -return words.

Lemma 24. Let $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$.

1. If \mathbf{u} has Property $G\text{-lps}(1)$, then \mathbf{u} has Property $G\text{-crw}(1)$.
2. If \mathbf{u} has Property $G\text{-lps}(N)$ with $N > 1$, then \mathbf{u} has Property $G\text{-crw}(N - 1)$.

Proof. We prove both claims simultaneously. If $N = 1$, set $M = 1$. Otherwise set $M = N - 1$. We prove by contradiction that \mathbf{u} satisfies Property $G\text{-crw}(M)$.

Suppose there is a factor $w \in \mathcal{L}(\mathbf{u})$, $|w| \geq M$, such that there is a factor $v \in \mathcal{L}(\mathbf{u})$ which is a complete G -return word of $[w]$ and is not a G -palindrome. Denote p the prefix of \mathbf{u} ending in the leftmost occurrence of a factor from $[v]$. It is clear that $|p| \geq N$.

Since p ends in a non-empty G -palindromic complete G -return word, the suffix of p of length 1 has at least two G -occurrences and thus Property $G\text{-lps}(N)$ assures that p

has a non-empty G -longest palindromic suffix which is G -unioccurrent. Let us denote $x := G\text{-lps}(p)$.

If $0 < |x| \leq |w|$, then x has at least two G -occurrences in p - a contradiction.

If $|w| < |x| < |v|$, then we can find a third G -occurrence of w in v - a contradiction with v being a complete G -return word of $[w]$.

If $|x| = |v|$, then we have a contradiction with v not being a G -palindrome.

If $|x| > |v|$, then we can find a factor $v' \in [v]$ such that its occurrence is in contradiction with the choice of the prefix p .

□

Remark 25. Again the assumptions of the previous lemma can be reduced, as it is visible in our proof. It is enough to require that any prefix v of \mathbf{u} of length greater than or equal to N has unique G -longest palindromic suffix or the last letter of the prefix v is G -unioccurrent in v .

6 G -defect

Richness in the classical sense is closely related to the notion of defect of a word. As introduced in [7], the *defect of a finite word* w is defined as follows

$$D(w) = |w| + 1 - \#\text{Pal}(w).$$

The *defect of an infinite word* \mathbf{u} is defined as

$$D(\mathbf{u}) = \sup_{w \in \mathcal{L}(\mathbf{u})} \{D(w)\}.$$

The relation with richness is clear: a word is rich if its defect is zero and it is almost rich if its defect is finite. Since the defect of a finite word depends only on its length and number of palindromes contained in it, it is straightforward to see that for any $a \in \mathcal{A}$ we have

$$D(w) \leq D(wa) \leq D(w) + 1, \quad D(w) \leq D(aw) \leq D(w) + 1, \quad \text{and} \quad D(w) = D(R(w)).$$

The question we address here is how to define a G -analogy of defect when the group G contains more than two elements. Of course, we would like to find a definition of G -defect such that G -richness and almost G -richness would be again connected with G -defect in a reasonable way.

Let us illustrate on the Thue-Morse word \mathbf{u}_{TM} the number of distinct G -palindromes contained in its factors. The Thue-Morse word $\mathbf{u} \in \{0, 1\}^{\mathbb{N}}$ is defined as the fixed point of the morphism $\sigma(0) = 01$ and $\sigma(1) = 10$ with the starting letter 0, i.e.,

$$\mathbf{u}_{TM} = 011010011001011010010110011010011001011001101001011010 \dots$$

The language of the Thue-Morse word is invariant under the reversal mapping R and under the antimorphism Θ which permutes letters 0 and 1. In [23], we showed that the Thue-Morse word is G -rich for $G = \{\text{Id}, R, \Theta, R\Theta\}$. In Table 1, the numbers of G -palindromic

| n | $\#\text{Pal}_R(p_n)$ | $\#\text{Pal}_\Theta(p_n)$ | $G\text{-lps}(p_n)$ | n | $\#\text{Pal}_R(p_n)$ | $\#\text{Pal}_\Theta(p_n)$ | $G\text{-lps}(p_n)$ |
|-----|-----------------------|----------------------------|---------------------|-----|-----------------------|----------------------------|---------------------|
| 0 | 1 | 1 | ε | 10 | 9 | 8 | 100110 |
| 1 | 2 | 1 | 0 | 11 | 10 | 9 | 001100 |
| 2 | 3 | 2 | 01 | 12 | 11 | 10 | 10011001 |
| 3 | 4 | 2 | 11 | 13 | 12 | 10 | 0100110010 |
| 4 | 5 | 3 | 0110 | 14 | 13 | 11 | 101001100101 |
| 5 | 6 | 3 | 101 | 15 | 14 | 12 | 11010011001011 |
| 6 | 7 | 4 | 1010 | 16 | 15 | 13 | 0110100110010110 |
| 7 | 8 | 5 | 110100 | 17 | 16 | 13 | 101101 |
| 8 | 9 | 6 | 01101001 | 18 | 17 | 13 | 01011010 |
| 9 | 9 | 7 | 0011 | 19 | 18 | 13 | 0010110100 |

Table 1: Count of palindromes and Θ -palindromes in the prefixes of the Thue-Morse word \mathbf{u}_{TM} . The prefix of \mathbf{u}_{TM} length n is denoted p_n .

factors of short prefixes of the Thue-Morse words are depicted. To generalize the notion of defect, the counting of G -palindromes must be replaced by counting the classes $[w]$ of G -palindromes. Thus, we define the set, denoted $\text{Pal}_G(w)$, of all G -palindromic classes of equivalence in a finite word w as follows

$$\text{Pal}_G(w) := \{[v] \mid v \text{ is a factor of } w \text{ and a } G\text{-palindrome}\}.$$

Definition 26. Let w be a finite word. The G -defect of w is defined as

$$D_G(w) := |w| + 1 - \#\text{Pal}_G(w) - \gamma_G(w),$$

where

$$\gamma_G(w) = \#\{[a] \mid \forall \Theta \in G^{(2)}, \Theta(a) \neq a, a \text{ is a factor of } w\}.$$

It follows from the definition that for all $w \in \mathcal{A}^*$ and $\mu \in G$ it holds $D_G(w) = D_G(\mu(w))$.

The authors of [15] also observed that the classical richness of w can be characterized by so-called Property *Ju*: *Any prefix of w has unioccurrent longest palindromic suffix.* The notion of the longest palindromic suffix helps to calculate the defect of a word. For a word w and a letter a , one can see that

$$D(wa) = \begin{cases} D(w), & \text{if } wa \text{ has unioccurrent longest palindromic suffix} \\ D(w) + 1, & \text{otherwise.} \end{cases}$$

Therefore, the defect of a finite word $w = w_1w_2 \dots w_n$ equals to the number of indices i for which $w_1w_2 \dots w_{i-1}w_i$ does not have a unioccurrent longest palindromic suffix. Such indices are in [6] called lacunas and in [16] defective positions. Inspired by this, we adopt the following definition of G -defective positions.

Definition 27. Let $w = w_1 \dots w_n \in \mathcal{A}$. An integer $1 \leq i \leq n$ is called *G-defective position* in w if w_i and $G\text{-lps}(w_1 \dots w_i)$ are not *G-unioccurrent* in $w_1 \dots w_i$. Otherwise, i is called a *G-non-defective position*.

The next lemma follows from comparing the last definition and the definition of *G-defect*.

Lemma 28. Let $w \in \mathcal{A}$, then

$$D_G(w) = \#\{1 \leq i \leq n \mid i \text{ is a } G\text{-defective position in } w\}.$$

Moreover, it can be easily shown that the following relations are preserved:

$$D_G(w) \leq D_G(wa) \leq D_G(w) + 1 \text{ and } D_G(w) \leq D_G(aw) \leq D_G(w) + 1$$

for all $w \in \mathcal{A}^*$ and $a \in \mathcal{A}$. Therefore, we can define *G-defect* of an infinite word.

Definition 29. Let \mathbf{u} be an infinite word. The *G-defect* of \mathbf{u} , denoted $D_G(\mathbf{u})$, is defined as

$$D_G(\mathbf{u}) := \sup_{w \in \mathcal{L}(\mathbf{u})} \{D_G(w)\}.$$

The immediate connection with Property $G\text{-lps}(N)$ is summarized in the following lemma.

Lemma 30. Let \mathbf{u} be an infinite word with language closed under G .

1. $D_G(\mathbf{u}) = 0$ if and only if \mathbf{u} satisfies Property $G\text{-lps}(1)$.
2. If there exists an integer N such that \mathbf{u} satisfies Property $G\text{-lps}(N)$, then $D_G(\mathbf{u})$ is finite.
3. If \mathbf{u} is uniformly recurrent and $D_G(\mathbf{u})$ is finite, then there exists an integer N such that \mathbf{u} satisfies Property $G\text{-lps}(N)$.

Proof. The first two claims follow from Lemma 28.

To show the last claim, suppose $D_G(\mathbf{u})$ to be finite. Then there exists a prefix $v := u_0 u_1 \dots u_{M-1}$ of \mathbf{u} such that $D_G(\mathbf{u}) = D_G(v)$ and any letter of the alphabet occurs in v . As \mathbf{u} is uniformly recurrent, there exists a constant N such that any factor w of \mathbf{u} of length at least $N - 1$ contains the prefix v as its factor. Using the definition and basic properties of *G-defect*, and maximality of $D_G(v)$, we obtain $D_G(v) = D_G(w) = D_G(wa)$ for any $a \in \mathcal{A}$ such that $wa \in \mathcal{L}(\mathbf{u})$. Therefore, the last position in wa is not *G-defective*, i.e., \mathbf{u} has Property $G\text{-lps}(N)$. □

It remains to connect *G-defect* with (almost) *G-richness*.

Theorem 31. Let \mathbf{u} be an infinite word with language closed under G .

1. $D_G(\mathbf{u}) = 0$ if and only if \mathbf{u} is G -rich.
2. If \mathbf{u} is uniformly recurrent, then $D_G(\mathbf{u})$ is finite if and only if \mathbf{u} is almost G -rich.

Proof. The first part follows from Lemma 30 and Theorem 20.

To show the second part, one can see that it follows from Lemma 30 that \mathbf{u} satisfies Property G -lps(N) for some N if and only if $D_G(\mathbf{u})$ is finite. We can then use Lemmas 22 and 24 to get equivalence with having Property G -crw(N') for some N' . Finally, we use Lemmas 16 and 19 to prove the equivalence with having Property G -graph(N') which is by definition equivalent with almost G -richness of \mathbf{u} . □

7 G -richness and bilateral order

As we have state in Theorem 7, point 7, words rich in classical sense can be characterized using bilateral order of bispecial factors. The proof of this fact given in [3] is based on the validity of point 5 of Theorem 7. In [23], a weaker form of G -analogy of this statement is given.

Proposition 32. *Let \mathbf{u} be an infinite word with language closed under G , M be a G -distinguishing number on \mathbf{u} , and N be an integer $N \geq M$. The word \mathbf{u} satisfies Property G -graph(N) if and only if*

$$\Delta\mathcal{C}(n) + \#G = \sum_{\Theta \in G^{(2)}} \left(\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n+1) \right) \quad \text{for any } n \geq N. \quad (7)$$

Unfortunately, we have no modification of the previous proposition describing Property G -graph(N) for N which are not G -distinguishing. Therefore, in this section we concentrate on the notion almost G -richness. Let us rephrase the previous proposition in a more handy way.

Corollary 33. *Let \mathbf{u} be an infinite word with language closed under G , M be a G -distinguishing number on \mathbf{u} , and N be an integer $N \geq M$. Then \mathbf{u} satisfies Property G -graph(N) if and only if*

1.

$$\Delta\mathcal{C}(N) + \#G = \sum_{\Theta \in G^{(2)}} \left(\mathcal{P}_\Theta(N) + \mathcal{P}_\Theta(N+1) \right)$$

2. and for all $n \geq N$, we have

$$\Delta^2\mathcal{C}(n) = \sum_{\Theta \in G^{(2)}} \sum_{\substack{w \in \mathcal{L}_n(\mathbf{u}) \\ w = \Theta(w)}} (\#\text{Pext}_\Theta(w) - 1).$$

Proof. The task to verify equalities $a(n) = b(n)$ for all $n \geq N$ means to verify $a(N) = b(N)$ and $\Delta a(n) = \Delta b(n)$ for all $n \geq N$. Let us consider $a(n)$ to be equal to the left-hand side and $b(n)$ to the right-hand side of (7). It is now enough to realize that

$$\Delta b(n) = \sum_{\Theta \in G^{(2)}} \mathcal{P}_{\Theta}(n+2) - \mathcal{P}_{\Theta}(n) = \sum_{\Theta \in G^{(2)}} \sum_{\substack{w \in \mathcal{L}_n(\mathbf{u}) \\ w = \Theta(w)}} (\#\text{Pext}_{\Theta}(w) - 1),$$

where we used equality (1). \square

Proposition 34. *Let $N \in \mathbb{N}$, $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ satisfy $G\text{-graph}(N)$, and w be a bispecial factor of \mathbf{u} of length at least N .*

- *If w is not a G -palindrome, then*

$$b(w) \geq 0.$$

- *If w is a Θ -palindrome for an antimorphism $\Theta \in G$, then*

$$b(w) \geq \#\text{Pext}_{\Theta}(w) - 1.$$

Proof. Let w be a bispecial factor having its length $M := |w| \geq N$ such that for all antimorphism $\Theta \in G$, $\Theta(w) \neq w$. By the definition of $b(w)$, we want to prove

$$\#\text{Bext}(w) \geq \#\text{Rext}(w) + \#\text{Lext}(w) - 1. \quad (8)$$

Let $B(w)$ be a bipartite graph with the set of vertices

$$V(w) = \{aw \mid a \in \text{Lext}(w)\} \cup \{wb \mid b \in \text{Rext}(w)\}.$$

There is an edge connecting vertices aw and wb if the word awb is a factor of \mathbf{u} . The number of vertices in the graph $B(w)$ is $\#\text{Rext}(w) + \#\text{Lext}(w)$ and the number of edges is $\#\text{Bext}(w)$. In the sequel, we show that this graph is connected. Since in any connected graph the number of edges equals at least the number of vertices minus one, the inequality (8) follows.

Let (k_n) be an increasing sequence of indices such that $k_0 > 0$ is an occurrence of w and k_n is an G -occurrence of w for any $n \geq 1$. Moreover, any G -occurrence of w in the suffix $u_{k_0}u_{k_0+1}u_{k_0+2}\dots$ of \mathbf{u} belongs to the sequence (k_n) . As \mathbf{u} satisfies Property $G\text{-graph}(N)$, then according to Remark 17 we have

$$\begin{aligned} u_{k_1}u_{k_1+1}\dots u_{k_1+M-1} &= \nu_1(w), \text{ where } \nu_1 \in G \text{ is an antimorphism,} \\ u_{k_2}u_{k_2+1}\dots u_{k_2+M-1} &= \nu_2(w), \text{ where } \nu_2 \in G \text{ is a morphism,} \\ u_{k_3}u_{k_3+1}\dots u_{k_3+M-1} &= \nu_3(w), \text{ where } \nu_3 \in G \text{ is an antimorphism,} \\ &\text{and so on.} \end{aligned}$$

The restriction of $\nu \in G$ to the set of letters is just a permutation. Therefore, for any

$\nu \in G$ and any $b \in \mathcal{A}$, there exists a letter $a \in \mathcal{A}$ such that $b = \nu(a)$. Thus, for any $n \in \mathbb{N}$, the factor $u_{k_n-1}u_{k_n}u_{k_n+1} \dots u_{k_n+M}$ can be written as

$$u_{k_n-1}u_{k_n}u_{k_n+1} \dots u_{k_n+M} = \nu_n(c_n)\nu_n(w)\nu_n(d_n) \quad \text{for some letters } c_n \text{ and } d_n.$$

As ν_{2i} is a morphism and ν_{2i+1} is an antimorphism, we have

$$c_{2i}wd_{2i} \in \mathcal{L}(\mathbf{u}) \quad \text{and} \quad d_{2i-1}wc_{2i-1} \in \mathcal{L}(\mathbf{u}). \quad (9)$$

Because w is not a G -palindrome, any G -occurrence of w together with the left and the right neighboring letters corresponds to a unique edge in the graph $B(w)$. For any $n \in \mathbb{N}$, the factor

$$\nu_n(w)\nu_n(d_n) \dots \nu_{n+1}(c_{n+1})\nu_{n+1}(w)$$

is a complete G -return word of $[w]$. According to Lemma 15, there exists an antimorphism $\Theta \in G$ such that $\nu_{n+1}(w) = \Theta(\nu_n(w))$ and $\nu_{n+1}(c_{n+1}) = \Theta(\nu_n(d_n))$. As $\nu_{n+1}(w) = \Theta(\nu_n(w))$ implies $\nu_{n+1} = \Theta\nu_n$, we get $\nu_{n+1}(c_{n+1}) = \Theta(\nu_n(c_{n+1})) = \Theta(\nu_n(d_n))$ and thus $c_{n+1} = d_n$. Using (9) we obtain

$$c_{2i}wc_{2i+1} \in \mathcal{L}(\mathbf{u}) \quad \text{and} \quad c_{2i}wc_{2i-1} \in \mathcal{L}(\mathbf{u}).$$

Recurrence of \mathbf{u} implies

$$V(w) = \{c_{2i}w \mid i \in \mathbb{N}\} \cup \{wc_{2i-1} \mid i \in \mathbb{N}\}.$$

Walking along \mathbf{u} , each G -occurrence of w represents an unordered edge connecting $c_{2i}w$ with wc_{2i-1} or $c_{2i}w$ with wc_{2i+1} , and thus in fact walking along \mathbf{u} represents a walk in the graph $B(w)$. Since any factor aw and wb occurs in $\mathcal{L}(\mathbf{u})$ infinitely many times, this walk in the graph $B(w)$ uses all vertices of $B(w)$. Therefore, the graph $B(w)$ is connected.

Now consider a G -palindromic bispecial factor w and denote by Θ the antimorphism such that $w = \Theta(w)$. We define the bipartite graph $B(w)$ in the same way as before. If $awb \in \mathcal{L}(\mathbf{u})$ and $b \neq \Theta(a)$, then $B(w)$ contains with the edge awb also the different edge $\Theta(b)w\Theta(a)$. Therefore, any G -occurrence of w together with the left and the right neighboring letters corresponds to a pair of edges in the graph $B(w)$ unless it represents a Θ -palindromic extension $aw\Theta(a)$. Let us replace the graph $B(w)$ by the graph $B'(w)$ in which vertices are couples $\{aw, w\Theta(a)\}$ and edges are either couples $\{awb, \Theta(b)w\Theta(a)\}$ or loops representing a Θ -palindromic extension $\{aw\Theta(a)\}$. Now we can interpret a walk along \mathbf{u} as a walk in the new graph $B'(w)$. Consequently, the graph $B'(w)$ must be connected. The connectivity of $B'(w)$ implies that the number of edges in $B'(w)$ which are not loops is at least equal to $\#\text{Rext}(w) - 1$. Since

$$\#\{awb \mid b \neq \Theta(a)\} = 2 \times \text{number of edges in } B'(w) \text{ which are not loops},$$

we obtain

$$\#\text{Bext}(w) = \#\{awb \mid b \neq \Theta(a)\} + \#\text{Pext}_\Theta(w) \geq 2(\#\text{Rext}(w) - 1) + \#\text{Pext}_\Theta(w)$$

As $\#Rext(w) = \#Lext(w)$, we deduce

$$b(w) = \#Bext(w) - \#Rext(w) - \#Lext(w) + 1 \geq \#Pext_{\Theta}(w) - 1.$$

□

Proposition 35. *Let \mathbf{u} be an infinite word with language closed under G , M be G -distinguishing on \mathbf{u} , and $N \geq M$ be an integer. Then \mathbf{u} has Property G -graph(N) if and only if any bispecial factor w of \mathbf{u} of length at least N satisfies:*

- if w is not a G -palindrome, then

$$b(w) = 0,$$

- if w is a Θ -palindrome for some $\Theta \in G$, then

$$b(w) = \#Pext_{\Theta}(w) - 1;$$

and

$$\Delta\mathcal{C}(N) + \#G = \sum_{\Theta \in G^{(2)}} \left(\mathcal{P}_{\Theta}(N) + \mathcal{P}_{\Theta}(N+1) \right).$$

Proof. (\Leftarrow): The assumption on bilateral orders and the fact that non-bispecial Θ -palindromic factors have a unique Θ -palindromic extension guarantee the following equality for all $n \geq N$:

$$\Delta^2\mathcal{C}(n) = \sum_{w \in \mathcal{L}_n(\mathbf{u})} b(w) = \sum_{\Theta \in G^{(2)}} \sum_{\substack{w \in \mathcal{L}_n(\mathbf{u}) \\ w = \Theta(w)}} (\#Pext_{\Theta}(w) - 1). \quad (10)$$

According to Corollary 33 it implies that \mathbf{u} satisfies Property G -graph(N).

(\Rightarrow): Let $n \geq N$. Using Proposition 34 and Corollary 33 we obtain

$$\begin{aligned} \Delta^2\mathcal{C}(n) &= \sum_{w \in \mathcal{L}_n(\mathbf{u})} b(w) = \sum_{\substack{w \in \mathcal{L}_n(\mathbf{u}) \\ w \neq \Theta(w) \\ \text{for all } \Theta}} b(w) + \sum_{\Theta \in G^{(2)}} \sum_{\substack{w \in \mathcal{L}_n(\mathbf{u}) \\ w = \Theta(w)}} b(w) \geq \\ &\geq \sum_{\Theta \in G^{(2)}} \sum_{\substack{w \in \mathcal{L}_n(\mathbf{u}) \\ w = \Theta(w)}} (\#Pext_{\Theta}(w) - 1) = \Delta^2\mathcal{C}(n) \end{aligned}$$

As the beginning and the end of our estimates is the same, the inequalities for $b(w)$ deduced in Proposition 34 must be equalities, which was to prove.

□

8 Examples

It is well known that any finite Coxeter group is isomorphic to a group generated by reflexion symmetries of a polygon in \mathbb{R}^d . The dihedral groups, usually denoted D_m , are groups of symmetries of the regular m -gon in the plane. For example, the regular pentagon depicted on Figure 1 has five reflexion symmetries of order 2 and four rotation symmetries of order 5 and identity of order 1, together D_5 has 10 elements. In general, the group D_m has $2m$ elements. As shown in [26], the generalized Thue-Morse word $t_{b,m}$ for all parameters $b \geq 2$ and $m \geq 1$, already considered in [24], is D_m -rich.

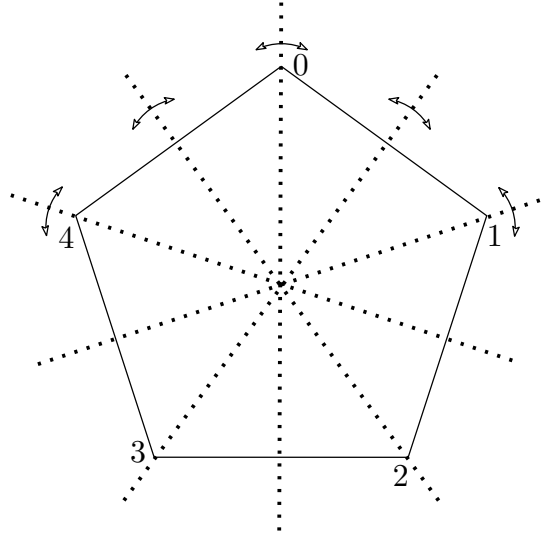


Figure 1: Regular pentagon and its symmetries.

In this section we provide 2 examples of G -rich words, where G is not a dihedral group. In the first example, the group is isomorphic to the group of symmetries of an orthogonal prism with rhomb base, see Figure 2a. In the second example, the group is isomorphic to the group of symmetries of two pyramids joint together by their rectangular bases, see Figure 2b.

The first group, denoted G , is constructed on an 8-letter alphabet $\mathcal{A} := \{0, 1, \dots, 7\}$. The antimorphisms Θ_0, Θ_1 and Θ_2 are defined on \mathcal{A}^* as follows

$$\begin{aligned}\Theta_0 : & 0 \mapsto 2, 1 \mapsto 1, 2 \mapsto 0, 3 \mapsto 3, 4 \mapsto 6, 5 \mapsto 5, 6 \mapsto 4, 7 \mapsto 7, \\ \Theta_1 : & 0 \mapsto 4, 1 \mapsto 5, 2 \mapsto 6, 3 \mapsto 7, 4 \mapsto 0, 5 \mapsto 1, 6 \mapsto 2, 7 \mapsto 3, \\ \Theta_2 : & 0 \mapsto 0, 1 \mapsto 3, 2 \mapsto 2, 3 \mapsto 1, 4 \mapsto 4, 5 \mapsto 7, 6 \mapsto 6, 7 \mapsto 5.\end{aligned}$$

The group $G \subset AM(\mathcal{A}^*)$ is the group generated by Θ_0, Θ_1 and Θ_2 .

The second group, denoted H , is on a 6-letter alphabet $\mathcal{B} := \{0, 1, \dots, 5\}$ and $H \subset$

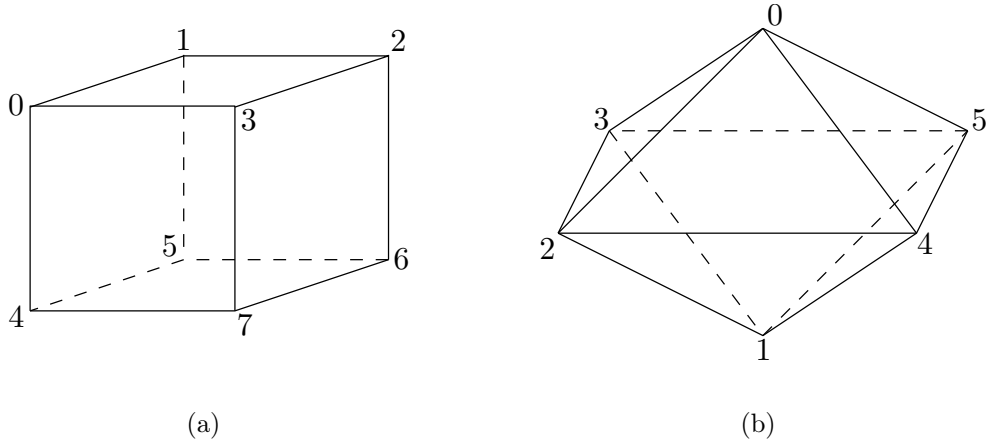


Figure 2: (a) Symmetries of $\mathcal{L}(\mathbf{u})$ from Example 36 represented by the symmetries of an orthogonal prism with rhomb base. (b) Symmetries of $\mathcal{L}(\mathbf{v})$ from Example 37 represented by two pyramids joint together by their rectangular bases.

$AM(\mathcal{B}^*)$ is generated by the 3 following antimorphisms:

$$\begin{aligned}\Psi_0 : & 0 \mapsto 0, 1 \mapsto 1, 2 \mapsto 4, 3 \mapsto 5, 4 \mapsto 2, 5 \mapsto 3, \\ \Psi_1 : & 0 \mapsto 1, 1 \mapsto 0, 2 \mapsto 2, 3 \mapsto 3, 4 \mapsto 4, 5 \mapsto 5, \\ \Psi_2 : & 0 \mapsto 0, 1 \mapsto 1, 2 \mapsto 3, 3 \mapsto 2, 4 \mapsto 5, 5 \mapsto 4.\end{aligned}$$

Example 36. Let $\varphi : \mathcal{A}^* \mapsto \mathcal{A}^*$ be a morphism defined as

$$\varphi : 0 \mapsto 01, 1 \mapsto 2, 2 \mapsto 65, 3 \mapsto 4, 4 \mapsto 23, 5 \mapsto 6, 6 \mapsto 47, 7 \mapsto 0.$$

Denote by \mathbf{u} the fixed point of φ . In Section 8.1, we show that \mathbf{u} has its language closed under G and \mathbf{u} is G -rich.

Example 37. Let $\mu : \mathcal{A}^* \mapsto \mathcal{B}^*$ be a morphism defined as

$$\mu : 0 \mapsto 15, 1 \mapsto 04, 2 \mapsto 12, 3 \mapsto 03, 4 \mapsto 04, 5 \mapsto 12, 6 \mapsto 03, 7 \mapsto 15.$$

Let $\mathbf{v} = \mu(\mathbf{u})$. In Section 8.2, we show that $\mathcal{L}(\mathbf{v})$ is closed under H and \mathbf{v} is H -rich.

The proofs of properties of \mathbf{u} and \mathbf{v} are split into several lemmas. Instead of their complete proofs, we provide just sketches or hints for readers.

8.1 G -richness of \mathbf{u}

To prove that \mathbf{u} defined in Example 36 is G -rich means to show that \mathbf{u} has Property G -graph(1). For this reason, we exploit Proposition 35. Therefore, one needs to study bispecial factors occurring in \mathbf{u} . We use a general method for circular morphisms described in [21]. In the case of φ , it yields simple results. To describe the bispecial factors of \mathbf{u} , we introduce the mapping $\pi : \{0, 2, 4, 6\} \mapsto \{0, 2, 4, 6\}$ as follows

$$\pi : 0 \mapsto 2, 4 \mapsto 0, 2 \mapsto 4, 6 \mapsto 6.$$

Lemma 38. *Let $w = w_0 \dots w_{n-1}$ be a non-empty bispecial factor of \mathbf{u} . Then $w_{n-1} \in \{0, 2, 4, 6\}$ and $\varphi(w)\pi(w_{n-1})$ is a bispecial factor of \mathbf{u} . Moreover, $b(w) = b(\varphi(w)\pi(w_{n-1}))$.*

Proof. The claim follows from the definition of φ and the fact that right special factors of length 1 are factors 0, 2, 4 and 6, and each has 2 right extensions. \square

The next statement can be easily deduced from the form of φ as well.

Lemma 39. *Let w , $|w| \geq 2$, be a bispecial factor of \mathbf{u} . Then there exists a unique bispecial factor of \mathbf{u} , say $v = v_0 v_1 \dots v_{m-1}$, such that*

$$w = \varphi(v)\pi(v_{m-1}). \quad (11)$$

According to the last two lemmas, all bispecial factors can be constructed from bispecial factors of length 1 using recursively formula (11). In fact, all the letters of \mathcal{A} are bispecial factors and G -palindromes. We show that the formula (11) produces from a G -palindrome again a G -palindrome.

Lemma 40. *For all $i \in \mathbb{Z}_3$ and $w \in \mathcal{L}(\mathbf{u})$, $w = w_0 \dots w_{n-1}$, we have*

$$x_{i-1}(w)\Theta_i\varphi(w) = \varphi\Theta_{i-1}(w)y_i(w),$$

where $y_i(w) := \Theta_i(\varphi(w_0))_0$ and $x_i(w) = (\varphi\Theta_i(w_{n-1}))_0$ and $(v)_0$ denotes the first letter of a word v .

Sketch of the proof. The proof is done by induction on n . Supposing the claim holds for n , one needs to deal with different cases according to the value i and possible factors $w_{n-1}w_n \in \mathcal{L}_2(\mathbf{u}) = \{54, 62, 47, 12, 04, 76, 65, 40, 01, 23, 30, 26\}$. The claim then follows from the definitions of φ , Θ_i and Θ_{i-1} . \square

Lemma 41. *Let $w = w_0 \dots w_{n-1}$ be a non-empty bispecial factor of \mathbf{u} . Then w is a Θ -palindrome, $\Theta \in G$, and $b(w) = \#\text{Pext}_\Theta(w) - 1$.*

Proof. As the bilateral order of a bispecial factor of length 1 equals 0, according to Lemmas 38 and 39, all bispecial factors have their bilateral order equal to 0. It is also clear that they have 2 right and 2 left extensions.

If $w = w_0 \dots w_{n-1}$ is a non-empty bispecial factor and if w is a Θ_{i-1} -palindrome for $i \in \mathbb{Z}_3$, one can show that $\pi(w_{n-1}) = y_i(w)$ and thus, $\varphi(w)\pi(w_{n-1})$ is a Θ_i -palindrome. Since the bispecial factors of length 1 are Θ_2 -palindromes, all bispecial factors are Θ_i -palindromes for some $i \in \mathbb{Z}_3$.

It follows that for all $i \in \mathbb{Z}_3$, $\mathcal{L}(\mathbf{u})$ contains infinitely many Θ_i -palindromes. Therefore, $\mathcal{L}(\mathbf{u})$ is closed under G .

Let w be a non-empty bispecial factor. Since $b(w) = 0$, w has 2 left and 2 right extensions, w is a Θ -palindrome for a unique $\Theta \in G$, and $\mathcal{L}(\mathbf{u})$ is closed under G , one can see that the number of Θ -palindromic extensions of w is 1. \square

Proof of G -richness of \mathbf{u} defined in Example 36. At first, we realize that the generators Θ_0, Θ_1 and Θ_2 of the group G guarantee the number 1 to be G -distinguishing on any infinite word over \mathcal{A} . Because of Lemma 41 and Proposition 35, it remains to verify that $\Delta\mathcal{C}(1) + \#G$ equals the number of all G -palindromes of length 1 and 2. One can readily see that $\Delta\mathcal{C}(1) = 4$, $\#G = 8$, the number of G -palindromes of length 1 is 8 and the number of G -palindromes of length 2 is 4. \square

8.2 H -richness of \mathbf{v}

The proof of H -richness of \mathbf{v} is very similar to the previous proof. In order to use Proposition 35, we explore the bilateral orders and H -palindromic extensions of bispecial factors of \mathbf{v} .

We define the morphism $\eta : \mathcal{A}^* \mapsto \mathcal{B}^*$ as

$$\eta : 0 \mapsto 041, 1 \mapsto 120, 2 \mapsto 031, 3 \mapsto 150, 4 \mapsto 150, 5 \mapsto 041, 6 \mapsto 120, 7 \mapsto 031.$$

Let $w = w_0 \dots w_{n-1} \in \mathcal{A}^*$ be a non-empty factor of \mathbf{u} . It follows from $\mathcal{L}_2(\mathbf{u})$ and the definition of μ that $\mu(w)\eta(w_{n-1})$ is a factor of \mathbf{v} . The following lemma summarizes the relation between the bispecial factors of \mathbf{v} and of \mathbf{u} .

Lemma 42. *Let $w \in \mathcal{L}(\mathbf{u})$, $w = w_0 \dots w_{n-1}$, be a non-empty bispecial factor. Then $\mu(w)\eta(w_{n-1})$ is a bispecial factor of \mathbf{v} .*

On the other hand, if $v \in \mathcal{L}(\mathbf{v})$, $|v| \geq 5$, is a bispecial factor of \mathbf{v} , then there exists a unique non-empty bispecial factor $w \in \mathcal{L}(\mathbf{u})$ such that $v = \mu(w)\eta(w_{n-1})$.

Lemma 43. *Let $i \in \mathbb{Z}_3$. If $w \in \mathcal{L}(\mathbf{u})$ is a non-empty Θ_i -palindrome, then the factor $\mu(w)\eta(w_{n-1}) \in \mathcal{L}(\mathbf{v})$ is a Ψ_i -palindrome.*

Sketch of the proof. We induce on the length of w . Fix $i \in \mathbb{Z}_3$. Suppose the claim holds for $w = w_0 \dots w_{n-1} = \Theta_i(w)$. Take $z \in \mathcal{A}$ such that $zw\Theta_i(z) \in \mathcal{L}(\mathbf{u})$. The proof follows from the definition of μ, η , and possible factors $zw_0 \in \mathcal{L}_2(\mathbf{u})$. \square

Proof of H -richness of \mathbf{v} defined in Example 37. According to the previous lemma, it is clear that $\mathcal{L}(\mathbf{v})$ is closed under H . The properties of $\mathcal{L}(\mathbf{u})$ also imply that all bispecial factors of \mathbf{v} of length greater than or equal to 5 have bilateral order 0 and one Θ -palindromic extension, where $\Theta \in H$ is the unique antimorphism fixing the bispecial factor. For shorter bispecial factors, of length greater than 1, this property needs to be verified by hand and is left to the reader.

Since 2 is an H -distinguishing number on \mathbf{v} , Proposition 35 requires to evaluate $\Delta\mathcal{C}(2)$, $\mathcal{P}_\Theta(2)$ and $\mathcal{P}_\Theta(3)$ for all involutive antimorphism $\Theta \in H$. It is easy to verify that $\Delta\mathcal{C}(2) = 4$, $\sum_{\Theta \in H^{(2)}} \mathcal{P}_\Theta(2) = 0$ and $\sum_{\Theta \in H^{(2)}} \mathcal{P}_\Theta(3) = 12$. Since $\#H = 8$, according to Proposition 35, \mathbf{v} satisfies Property H -graph(2).

To claim that \mathbf{v} is H -rich, we need to verify that \mathbf{v} satisfies H -graph(1). Thus, it remains to show that all loops in $\Gamma_1(\mathbf{v})$ are H -palindromes and the graph obtained from $\Gamma_1(\mathbf{v})$ by removing loops is a tree. Since it can be easily verified by hand, the word \mathbf{v} is H -rich. \square

Denote for all $i \in \mathbb{Z}_3$ by H_i the subgroup of H generated by Ψ_i and $\Psi_{(i+1 \bmod 3)}$. It is easy to verify that $\#H_i = 4$ for all i , the number 1 is H_0 -distinguishing and H_1 -distinguishing, and the number 2 is H_2 -distinguishing. It follows from the last proof that $\mathcal{P}_{\Psi_i}(3) = 4$ and $\mathcal{P}_{\Psi_i}(2) = 0$ for all $i \in \mathbb{Z}_3$. One can also verify that $\mathcal{P}_{\Psi_0}(1) = \mathcal{P}_{\Psi_2}(1) = 2$ and $\mathcal{P}_{\Psi_1}(1) = 4$. Since $\Delta\mathcal{C}(1) = 2$, the word \mathbf{v} is H_0 -rich, H_1 -rich. Since $\Delta\mathcal{C}(2) = 4$, it is almost H_2 -rich (it is not H_2 -rich).

9 Comments and open problems

- The dihedral groups D_m form a special class of finite Coxeter groups. As shown in [26], for any dihedral group there exists a D_m -rich word. Is it possible for any given finite Coxeter group G to find a G -rich word?
- The list of examples of G -rich words is very short, the list of almost G -rich words (which are not G -rich) is empty. In [16], Glen et al. described a class of morphisms such that morphic image of a rich word under this morphism has a finite non-zero defect. Find a class of morphisms producing almost G -rich words with finite non-zero defect by applying this morphism to a G -rich word.
- For a word \mathbf{u} with language closed under reversal, Brlek and Reutenauer conjectured in [8] for the defect $D(\mathbf{u})$ that

$$2D(\mathbf{u}) = \sum_{n \in \mathbb{N}} T(n), \quad \text{where } T(n) := \Delta\mathcal{C}(n) + 2 - \mathcal{P}(n+1) - \mathcal{P}(n).$$

This equality was shown to be valid for periodic words in [8] and for uniformly recurrent words in [5].

Can the G -defect of an infinite word \mathbf{u} be expressed using the differences between right-hand and left-hand sides of inequalities in (3)?

- In Section 2.4, definitions of rich words and Θ -rich words were reminded. In our new terminology, they are $\{\text{Id}, R\}$ -rich words and $\{\text{Id}, \Theta\}$ -rich words respectively. The groups $\{\text{Id}, R\}$ and $\{\text{Id}, \Theta\}$ are clearly isomorphic. In [9], it is shown that a so-called Θ -standard word with seed, which is almost Θ -rich, is a morphic image of a standard Arnoux-Rauzy words, which is rich. In [22], we have a more general case: any uniformly recurrent almost Θ -rich word is a morphic image of a rich word. Is an almost G_1 -rich word related to a G_2 -rich word for some group G_2 isomorphic to G_1 ?
- Let \mathbf{u} be an infinite word having language closed under a group G . The closedness under G can be exploited to estimate the number of distinct frequencies of factors of the same length n . In [2], an upper bound on this number is given (for n being a G -distinguishing number). The estimate is based on the inequality from Theorem 3. Looking at the proof of the estimate, it can be seen that the only candidates

for reaching the upper bound for all sufficiently large n are almost G -rich words. However, as noted in [2], in our words, almost G -richness does not imply the upper bound to hold for all sufficiently large n .

- The Thue-Morse word is G -rich, where G is generated by two commuting antimorphisms R and Θ . However, it is not G_1 -rich while taking a proper subgroup G_1 of G . In our considerations, we did not assume the group G to be the maximal group of symmetries such that an infinite word \mathbf{u} is closed under G .

Suppose \mathbf{u} is an infinite word having language closed under a group G . Let G_1 be a proper subgroup of G containing at least one antimorphism. Suppose \mathbf{u} is both almost G -rich and almost G_1 -rich. Let N be a G -distinguishing number. Then, according to Proposition 32, we can for all $n \geq N$ write

$$\begin{aligned}\Delta\mathcal{C}(n) + \#G &= \sum_{\Theta \in G^{(2)}} \left(\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n+1) \right) \quad \text{and} \\ \Delta\mathcal{C}(n) + \#G_1 &= \sum_{\Theta \in G_1^{(2)}} \left(\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n+1) \right).\end{aligned}$$

Thus, we get

$$\#G - \#G_1 = \sum_{\Theta \in G^{(2)} \setminus G_1^{(2)}} \left(\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n+1) \right). \quad (12)$$

Since G_1 is a proper subgroup of G , we have $\#G = \ell \#G_1$ for $\ell > 1$. Take $\Theta \in G^{(2)} \setminus G_1^{(2)}$ and $w \in \mathcal{L}(\mathbf{u})$, $|w| \geq N$, such that $\Theta(w) = w$. Then for all $v \in [w]$, it can be show that there exists $\Psi \in G^{(2)} \setminus G_1^{(2)}$ such that v is a Ψ -palindrome. Since $\#[w] = \frac{\#G}{2}$, the right-hand side of (12) equals $k(n) \frac{\#G}{2}$ for $k(n) \in \mathbb{N}$. We get from (12) that

$$(\ell - 1)\#G_1 = k(n)\ell \frac{\#G_1}{2}.$$

The only solution is $\ell = 2$ and $k(n) = 1$ for all n . Thus, we obtain the following condition

$$\#G_1 = \frac{1}{2}\#G = \sum_{\Theta \in G^{(2)} \setminus G_1^{(2)}} \left(\mathcal{P}_\Theta(n) + \mathcal{P}_\Theta(n+1) \right) \quad \text{for all } n \geq N.$$

Indeed, these conditions are satisfied for the three subgroups H_0 , H_1 and H_2 of the group H and the word \mathbf{v} , see the last part of Section 8.2.

Further characterization of such group and examples of such infinite words is an open problem.

Acknowledgments

This work was supported by the Czech Science Foundation grant GAČR 201/09/0584, by the grants MSM6840770039 and LC06002 of the Ministry of Education, Youth, and Sports of the Czech Republic, and by the grant of the Grant Agency of the Czech Technical University in Prague grant No. SGS11/162/OHK4/3T/14.

References

- [1] P. BALÁŽI, Z. MASÁKOVÁ, AND E. PELANTOVÁ, *Factor versus palindromic complexity of uniformly recurrent infinite words*, Theoret. Comput. Sci., 380 (2007), pp. 266–275.
- [2] L. BALKOVÁ, *Factor frequencies in languages invariant under more symmetries*, preprint available at <http://arxiv.org/abs/1107.0471>, (2011).
- [3] L. BALKOVÁ, E. PELANTOVÁ, AND Š. STAROSTA, *Sturmian jungle (or garden?) on multiliteral alphabets*, RAIRO-Theoret. Inf. Appl., 44 (2010), pp. 443–470.
- [4] —, *Infinite words with finite defect*, Adv. in Appl. Math., 47 (2011), pp. 562–574.
- [5] —, *On Brlek-Reutenauer conjecture*, Theoret. Comput. Sci., 412 (2011), pp. 5649–5655.
- [6] A. BLONDIN MASSÉ, S. BRLEK, A. GARON, AND S. LABBÉ, *Combinatorial properties of f -palindromes in the Thue-Morse sequence*, Pure Math. Appl., 19 (2008), pp. 39–52.
- [7] S. BRLEK, S. HAMEL, M. NIVAT, AND C. REUTENAUER, *On the palindromic complexity of infinite words*, Internat. J. Found. Comput., 15 (2004), pp. 293–306.
- [8] S. BRLEK AND C. REUTENAUER, *Complexity and palindromic defect of infinite words*, Theoret. Comput. Sci., 412 (2011), pp. 493–497.
- [9] M. BUCCI AND A. DE LUCA, *On a family of morphic images of arnoux-rauzy words*, in LATA '09: Proceedings of the 3rd International Conference on Language and Automata Theory and Applications, Berlin, Heidelberg, 2009, Springer-Verlag, pp. 259–266.
- [10] M. BUCCI, A. DE LUCA, A. GLEN, AND L. Q. ZAMBONI, *A connection between palindromic and factor complexity using return words*, Adv. in Appl. Math., 42 (2009), pp. 60–74.
- [11] —, *A new characteristic property of rich words*, Theoret. Comput. Sci., 410 (2009), pp. 2860–2863.

- [12] J. CASSAIGNE, *Complexity and special factors*, Bull. Belg. Math. Soc. Simon Stevin 4, 1 (1997), pp. 67–88.
- [13] D. DAMANIK AND D. LENZ, *Uniform spectral properties of one-dimensional quasicrystals, I. Absence of eigenvalues*, Commun. Math. Phys., 207 (1999), pp. 687–696.
- [14] A. DE LUCA AND A. DE LUCA, *Pseudopalindrome closure operators in free monoids*, Theoret. Comput. Sci., 362 (2006), pp. 282–300.
- [15] X. DROUBAY, J. JUSTIN, AND G. PIRILLO, *Episturmian words and some constructions of de Luca and Rauzy*, Theoret. Comput. Sci., 255 (2001), pp. 539–553.
- [16] A. GLEN, J. JUSTIN, S. WIDMER, AND L. Q. ZAMBONI, *Palindromic richness*, European J. Combin., 30 (2009), pp. 510–531.
- [17] A. HOF, O. KNILL, AND B. SIMON, *Singular continuous spectrum for palindromic Schrödinger operators*, Comm. Math. Phys., 174 (1995), pp. 149–159.
- [18] J. E. HUMPHREYS, *Reflection groups and Coxeter groups*, vol. 29 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1990.
- [19] L. KARI AND K. MAHALINGAM, *Watson-Crick conjugate and commutative words*, in DNA Computing, M. Garzon and H. Yan, eds., vol. 4848 of Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, 2008, pp. 273–283.
- [20] L. KARI AND K. MAHALINGAM, *Watson-Crick palindromes in DNA computing*, Nat. Comput., 9 (2010), pp. 297–316.
- [21] K. KLOUDA, *Non-standard numerations systems and combinatorics on words*, PhD thesis, Czech Technical University in Prague, 2010.
- [22] E. PELANTOVÁ AND Š. STAROSTA, *Infinite words rich and almost rich in generalized palindromes*, in Developments in Language Theory, G. Mauri and A. Leporati, eds., vol. 6795 of Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 406–416.
- [23] —, *Languages invariant under more symmetries: overlapping factors versus palindromic richness*, preprint available at <http://arxiv.org/abs/1103.4051>, (2011).
- [24] E. PROUHET, *Mémoire sur quelques relations entre les puissances des nombres*, C. R. Acad. Sci. Paris, 33 (1851), p. 225.
- [25] A. RESTIVO AND G. ROSONE, *Balancing and clustering of words in the Burrows-Wheeler transform*, Theoret. Comput. Sci., 412 (2011), pp. 3019–3032.
- [26] Š. STAROSTA, *Generalized Thue-Morse words and palindromic richness*, preprint available at <http://arxiv.org/abs/1104.2476>, (2011).

- [27] —, *On theta-palindromic richness*, Theoret. Comput. Sci., 412 (2011), pp. 1111–1121.